DOCUMENT RESUME

ED 269 472 TM 860 329

AUTHOR Banta, Trudy W.; Schneider, Janet A.
TITLE Using Locally Developed Comprehensive Exams for Majors to Assess and Improve Academic Program Quality.
PUB DATE 19 Apr 86
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (70th, San Francisco, CA, April 16-20, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Descriptive (141)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *College Curriculum; Curriculum Development; Curriculum Evaluation; Departments; Educational Quality; Higher Education; Item Analysis; Majors (Students); Outcomes of Education; *Program Evaluation; Program Improvement; Relevance (Education); Scoring; *Teacher Made Tests; *Test Construction; Testing
IDENTIFIERS *Comprehensive Examinations; *Exit Examinations; *University of Tennessee Knoxville

ABSTRACT

This paper briefly presents a history of comprehensive examinations in higher education in the United States, characteristics of comprehensive exams, and the use of such exams in assessing program quality. This paper then focuses on the experience of 11 program faculties at the University of Tennessee (Knoxville), that developed and used exit exams for majors as one component of a comprehensive program evaluation process. The purpose for test development is discussed, as well as test development procedures: (1) faculty involvement; (2) initial steps; (3) item types; (4) test length; (5) consultants; and (6) test revisions. The preparation of students, test administration and scoring are described, as is the post-administration evaluation of the tests through student reaction, faculty reaction, item analysis, test revision, and test security. The use of test results, future test use, and faculty reaction to the test development process are presented. Faculty involvement in the test development and a review of students' performance stimulated a variety of improvements, including increases in curricular structure, more consistency among faculty teaching core courses, stronger linkages between lower- and upper-division coursework, and more opportunities for students to apply knowledge learned in classes. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED269472

Using Locally Developed Comprehensive Exams for Majors To Assess and Improve Academic Program Quality

Trudy W. Banta
Research Professor

Janet A. Schneider
Graduate Student

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

T.W. Banta

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the Annual Meeting
of the
American Educational Research Association
San Francisco
April 19, 1986
Session 39.31

LEARNING RESEARCH CENTER
The University of Tennessee, Knoxville
1819 Andy Holt Avenue
Knoxville, Tennessee 37996-4350

TM 860 329

ABSTRACT

Traditionally, comprehensive exams in higher education have been used to assess levels of attainment of individual students. The growing emphasis on assessing quality in higher education encourages use of comprehensive exams to identify strengths and weaknesses of academic programs. At The University of Tennessee, Knoxville some 40 departments are using locally developed exit exams for majors as one component of a comprehensive program evaluation process. This paper summarizes the experience of eleven program faculties in developing and using such exams. Faculty involvement in test development and review of students' performance has stimulated a variety of improvements, including increases in curricular structure, more consistency among faculty in teaching core courses, stronger linkages between lower- and upper-division coursework, and more opportunities for students to apply knowledge learned in classes.

Using Locally Developed Comprehensive Exams for Majors To Assess and Improve Academic Program Quality

Trudy W. Banta Janet A. Schneider

As is the case in the history of many educational practices, interest in the use of comprehensive examinations in the undergraduate major field has waxed and waned. A review of the literature on testing indicates that the practice of examining students at various stages of their academic careers to assess the extent of their learning began in the earliest years of education in America. Comprehensive testing experienced a decline in use during the 1890s following the introduction of the elective system. In 1913 such testing began to enjoy a revival, but an apparent peak of interest in 1959 was not sustained through the 1970s. The current national interest in assessment of the outcomes of higher education has occasioned yet another increase in the use of comprehensive exams--this time for the purpose of providing evidence of the quality of educational programs rather than the level of individual student attainment.

History of Comprehensive Examinations in the United States

The colonial colleges administered annual public recitations that amounted to a test of rote memorization of factual content rehearsed in daily recitations over the course of the school year (Rudolph, 1978). "Comprehensive" at that time meant that the student was held responsible for any material that had been presented during the past year. In 1824 the University of Virginia was perhaps the first institution to require passage of general exams at the end of the student's chosen course of study (Levine, 1978). In the 1830s Yale introduced written exams at the end of both sophomore and senior years that enabled faculty to assess students' skills in written expression. Using the same questions as in the sophomore year to test each candidate for the degree permitted comparison and standard-setting for a class of students (Smallwood, 1935; Rudolph, 1978).

Until the late 19th century, academic courses of study were largely prescribed by the faculty; students shared a common learning experience and thus could be tested with common instruments in the areas of classical literature, philosophy and mathematics. The practice of giving comprehensive examinations declined during the 1890s when the elective system began to take root in an increasing number of colleges and universities following its introduction by President Eliot of Harvard College (Jones, 1933).

According to Jones, in 1913 Whitman College in Washington became the first institution to require all candidates for graduation to pass "an (oral) examination on the entire work of their major study (1933, p. 72)." Then-president Stephen Penrose acknowledged having been influenced by European methods and the reasonableness of the expectation that a graduate of the college should know enough about one field to express that knowledge adequately. In 1913 candidates for graduation from Harvard's division of history, government and economics were required to pass an exam that covered material relevant to the field of concentration though not necessarily addressed in the courses of study. By 1919 the exam became optional in all departments, with honors students being recognized for superior performance.

Surveys of the practice of administering comprehensive exams have varied in the range of the population covered (e.g., only liberal arts colleges, private religiously affiliated institutions, Carnegie Council institutional categories), and in the approaches used (e.g., letter surveys administered to presidents, content analyses of college catalogs). The authors have tended not to differentiate types of comprehensive examination practices in their reporting, e.g., general comprehensives, comprehensive exams in the major, senior comps administered in at least one department or to honors students only. While none of the reported surveys uses a sample that may be characterized as entirely representative of the population, a summary of the findings of all of them may provide a rough idea of the history of the practice of giving comprehensive exams.

Jones presented three decades of data that revealed an increase during the late 1920s and early 1930s in the use of comprehensive exams. By 1932, 13 percent of 654 colleges and universities used senior comprehensives in at least one department. In 1957, 33 percent of a sample of liberal arts institutions provided for comprehensive exams (Dressel & DeLisle, 1969). Of the 466 liberal arts institutions responding to Dressel and DeLisle's 1959 study, 52 percent (243) used some kind of senior comprehensive, most commonly to designate honors students (Dressel & Associates, 1961). A study using 1967 college catalogs revealed that 40 percent of liberal arts colleges and universities made provision for comprehensive exams (Dressel & DeLisle, 1969).

Singletary found that 310 (33%) of the 946 institutions in a survey that he reported in 1968 used some form of senior comprehensive exam for a selected group of students. Almost all of the small liberal arts colleges in the sample used such tests. By 1975 a Carnegie Council review of 270 college catalogs by Carnegie institutional categories showed that only 24 percent were using comprehensive exams (Levine, 1978, p. 90). These studies, which were conducted separately and independently, in chronological order, suggest marked fluctuations in the extent of use of senior comprehensives on the part of American colleges and universities.

These survey reports devote little attention to the reasons for the ebb and flow of interest in comprehensive exams. The upward trend noted by Jones in 1933 apparently was due to the advent of designating a major field of concentration, which began early in this century in a few colleges, including Harvard, and spread to most other colleges rather quickly. In 1909 Harvard required that students take a sufficient number of courses in one field of study to acquire depth of understanding in that field, and courses in all major branches of knowledge to acquire a broad understanding of all fields. Senior exams covered comprehensive knowledge of both general education and the field of concentration. The decrease in use of exams for graduating seniors in the 1960s that was noted by Singletary in his survey of public and private colleges and universities may be attributed to faculty reaction to student demands for more control of the curriculum and a growing skepticism about the validity of tests of all kinds. One of the few generalizations that can be drawn from the literature is that comprehensive testing has always been more prevalent in private liberal arts colleges than in other types of institutions.

Table 1

HISTORICAL PATTERN OF SENIOR COMPREHENSIVE TESTING IN THE UNITED STATES

Source (publication date)	Year of data collection	Percent using comp exams at senior level	Number of responding institutions
Jones (1933)	1933	13%	654 accredited institutions
Dressel & DeLisle (1969)	1957	33%	322 4-year liberal arts colleges
Dressel & Associates (1961)	1959	52%	466 liberal arts colleges
Singletary (1968)	1968	33%	946 liberal arts institutions
Dressel & DeLisle (1969)	1969	40%	322 4-year liberal arts colleges
Levine (1978)	1975	24%	270 colleges and universities

NOTE: While each study used a different sampling frame and different approaches for gathering data, these independently conducted surveys suggest an ebb and flow of interest in comprehensive testing at the senior level.

Characteristics of Comprehensive Exams

A variety of purposes has been suggested for the use of comprehensive examinations. Rudolph (1978) indicated that the rationale for examining at Harvard in 1919 was to provide "an instrument for bringing coherence and design and some semblance of unity to the academic course (p. 236)." Rudolph added that the comprehensive exam demonstrated that the university was serious about the curriculum, and students thus were encouraged to be serious about meeting its objectives. The objectives guiding Denison University in 1934 were "to measure the student's ability to correlate his knowledge effectively," both in command of the facts and principles in the field of concentration and in the ability to use this knowledge in new situations (Gordon, 1958, p. 622).

While the primary purpose of comprehensive exams throughout their long history in higher education has been to assess the levels of learning attained by individual students, in the 1980s there is a growing belief that such tests can be used to assess program quality. That is, the performances of individual students can be combined statistically, and relative strengths and weaknesses of curricula and instruction may be determined by studying mean scores and dispersion of scores on subparts of the exam.

Traditionally, comprehensive exams have been oral or written, dealing with the entire college curriculum or the major field of study, or both. Multiple-choice tests are rarely mentioned in the literature. Major comprehensives differ in length, form, coverage, and duration, and sometimes are accompanied by exams in the minor field of emphasis. At the time of the survey conducted by the Carnegie Council (Levine, 1978), a senior thesis or project was used more commonly than the comprehensive exam, with 41 percent of liberal arts colleges making use of one or both of these methods of evaluation.

Essay and objective tests are available commercially in some fields, but faculty-developed tests are an alternative that may be used along with a standardized test or alone. Dressel (1976) identified several issues connected with the various kinds of tests in common usage. On the matter of choosing between nationally standardized and locally developed tests, the author noted positive features of each. Standardized tests are technically superior, save faculty time, and provide norms for comparing scores with those of other institutions. Faculty-developed tests reflect the local curriculum more accurately, and are more likely to have faculty support than do standardized exams. Standardized tests theoretically resolve the question of credibility for purposes of gaining accreditation and establishing accountability, whereas the validity of tests designed by departmental faculty for these purposes can be called into question. However, locally developed tests also may be rigorously designed instruments for program evaluation when appropriate steps are taken to avoid threats to reliability and validity.

Dressel (1976) identified a number of factors to be considered in selecting the format of a comprehensive test to be developed locally. So-called "objective" tests (containing multiple-choice and true/false items) can cause faculty dissatisfaction because the coverage of content and skills is necessarily limited. Moreover, the faculty time involved in test

construction is costly, and lack of technical expertise may jeopardize reliability and validity. Concerns about test security require that new items be written periodically.

Dressel cautions that essay exams are deceptively easier to formulate than objective tests. Careful guidelines for scoring essays must be established and enforced. Oral examinations most commonly used with honors students in small, selective colleges require the participation of external examiners for development and grading. Dressel refers to the problem of submitting oral questions for advance approval, which diminishes the major strength of the oral format--that of interactive dialogue in which questions may be asked in response to the candidate's ongoing performance. On the other hand, prescreening is helpful in eliminating unreasonable questions.

Lack of a standard for comparison with other colleges is the most serious limitation of the locally developed test. However, when carefully done, exams designed by faculty usually provide a more accurate measure of student attainment of local objectives than do nationally standardized instruments. Clearly, the locally developed test is superior for evaluating the effectiveness of a given program.

A look at the history of America's on-again-off-again affair with comprehensive examinations in the undergraduate major reveals many of the same concerns that we have today. In 1933, Jones cited two prominent reasons for educators' aversion to examining of all kinds: (1) exams are "primarily hashed-over textbook items and . . . do not sample enough data to insure mastery of the material; " and (2) "examining is artificial"--exams that merely assess mastery of factual content neglect other valuable aspects of the curriculum (p. 15). Other scholars have noted other considerations: Harvard president A. L. Lowell cited the rising popularity of the elective system as contributing to the perceived lack of need or relevance for comprehensive exams, "each course being ended, closed and forever completed by its own exam (1912, pp.585-86);" faculty often resent the time required to prepare, administer, and evaluate exams; and finally, students may lack sufficient experience and preparation for the task of performing adequately on an integrative test of accumulated knowledge. These and other issues continue to trouble the academic community and, while institutions have devised their own individual methods for dealing with them, there remains a need for the development of a workable, valid system for providing effective and productive assessments of academic programs and student performance.

Use of Comprehensive Exams in Assessing Program Quality

Tennessee has become the first state to provide a portion of the funding for all public higher education institutions on the basis of the efforts of those institutions to use achievement tests and surveys to evaluate and improve their academic programs. A financial supplement of an amount equal to as much as five percent of each institution's education and general budget for instruction is awarded annually to institutions that test students in general education and the major and use the results of surveys of client groups to (1) establish the status of programs in meeting student development objectives, and (2) make program improvements as evaluation data warrant (Banta, 1985; and Banta & Fisher, 1984).

The attractiveness of the financial supplement has motivated departmental

faculty at the University of Tennessee, Knoxville--the state's public research institution with an undergraduate enrollment of approximately 20,00 and a graduate enrollment of 5,500--to select or develop comprehensive exams for use in the assessment of academic program quality. Approximately half of the academic programs have access to nationally standardized tests in the major, and most of these departments have elected to use the national exams. However, for the other departments no such test is available, so approximately 40 departmental faculties have elected to develop their own exam in the major.

The remainder of this paper will focus on the experience of eleven departments at UTK that constructed and administered an exam during 1983-84 or 1984-85, and thus have had an opportunity to make changes in curriculum and/or instruction on the basis of this experience. The eleven departments represent five colleges, and the group includes eight programs for majors at the baccalaureate level and three programs at the master's degree level. The degree programs for which comprehensive exams have been developed are:

College	Program Title	Degree Level Tested
Agriculture	Animal Science	BS
Agriculture	Food Technology & Science	BS
Agriculture	Ornamental Horticulture & Landscape Design	BS
Communications	Advertising	BS
Communications	Communication	MS
Education	Dance	BA
Education	Adult Education	MS
Human Ecology	Nutrition	MS
Human Ecology	Nutrition & Food Sciences	BS
Human Ecology	Textiles & Clothing	BS
Liberal Arts	Geography	BA

Purpose for Test Development

Almost all of the eleven departments that developed an exam were motivated to do so initially by the promise of a financial supplement to the University. However, once the dean of the college and the department head had committed the department to constructing a test, most faculties formulated their own rationale for proceeding with the task. Four of the departmental faculties envisaged a test that would provide an indication of the extent to which student majors were achieving the faculty's objectives for their skill and knowledge development in the field of study. Four departments also indicated that they would like information about the effectiveness of their teaching. Other reasons mentioned by a single department include:

- to assess the need for a core curriculum in the major field.
- to provide a way of reviewing course objectives as a component of the self-study for an academic program review.
- to provide a way to obtain faculty agreement on curriculum and instructional objectives.
- to provide leadership for the development of a certification exam for a national professional association.

- to improve the comprehensive exam for master's level students by adding a common core of items to which all students would respond.

Test Development Procedures

Faculty involvement in test development. All eleven departmental faculties decided to focus the comprehensive exam upon a core of common course work or objectives that they felt all majors should have mastered. The department head assumed leadership for test development, and in six of the units the head appointed a committee of three or four faculty to coordinate the process. In three departments the whole faculty worked on test items; in one unit a single individual wrote all the items. In Ornamental Horticulture all faculty had to agree on all items to be included in the exam. In most other departments, item-writing was delegated to subgroups of faculty by specialty area, and at some point in the process each faculty member had an opportunity to review the entire exam as put together by the coordinator(s).

First steps. All faculties began the test development process by defining the content areas to be included in the test. Two units in the Department of Nutrition and Food Sciences were able to start with sets of core competencies that had been formulated two years earlier within the department. Another unit worked from objectives for individual courses. However, in the majority of the departments, faculty simply began by generating questions within specific content areas. Six departments had a head-start on item development since they had access to one or more of the following:

- their own final exam file,
- a set of comprehensive exams from other universities,
- an item pool generated previously to test core competencies, or
- questions from diagnostic or placement exams administered to graduate students.

The key question that guided the work of most faculty was, "What should all students know when they finish the course work for a major?"

Types of items used. Essay questions were used exclusively by only two units--both testing students at the master's level. Five departments used the multiple-choice item format exclusively. All others utilized a combination of multiple-choice and other types of items, including essays, matching, short answers, and true/false items. These so-called "objective" items were selected because they were relatively easy to score, and provided maximum coverage of content. Faculty members recognized that most of the items they generated required the student to utilize only the simplest cognitive skills--recall and comprehension of information in the discipline--and they felt a need to develop questions that would test higher-order intellectual abilities such as problem-solving, analysis, synthesis, and evaluation. However, they found it very hard to develop the more complex items. The three units in the College of Human Ecology employed Bloom's taxonomy of educational objectives to classify each item, and their goal was to include items from each level of the taxonomy. But even these faculties had great difficulty generating items to test the highest levels of cognitive ability.

Length of test. There was a good deal of variation in the number of items used in the departmental exams and in the amount of time the faculty thought students should spend taking the tests. Three programs set 100 multiple-choice items as the limit, while two others included more than 200 items. Departments employing essays obviously had fewer items. Typically, the exams were scheduled for 60 to 90 minutes; the full range of test-taking times is listed below:

45 - 60 Minutes	(3 Programs)
1½ - 2 Hours	(4 Programs)
2 Hours	(2 Programs)
2½ Hours	(1 Program)
14 Hours (Take-home essays)	(1 Program)

Consultants. Every locally developed exam was required to be reviewed by two consultants outside the department. The faculties had the choice of using two specialists in the discipline, or one in the discipline and a measurement specialist. Eight of the eleven programs used two off-campus subject matter experts; three others utilized a measurement consultant on campus and an off-campus specialist in the field as evaluators. Five faculties looked for consultants among the faculty of programs at institutions they considered similar to their own. Others selected individuals whom they felt would take their work seriously and provide helpful feedback about the exam. Two departments specified that their consultants be involved in teaching and advising; three programs preferred individuals involved in research, and two others wanted specialists in the discipline noted for their evaluation skills.

The measurement consultants helped establish the clarity and quality of items, while specialists in the discipline validated the content of the exam. The consultants in the discipline often were furnished with a set of departmental objectives or competencies for students and asked to verify that the exam provided reasonable coverage of these objectives. For the faculties using Bloom's taxonomy, the disciplinary consultants also classified each item according to the level of cognitive ability they perceived it to measure. Six program faculties asked their consultants to pilot-test the UTK instrument with their students, while three other departments used samples of their own students as the pilot group. Three programs used only external consultant review in assessing item quality, i.e., there was no pilot test.

Kinds of revisions. As a result of pilot-testing and/or consultant review, faculties improved ambiguous items and usually shortened the instrument. The Dance faculty had included a performance measure, but their consultant in dance recommended that this one-time performance assessment not be used. (The faculty currently assess the progress of each student at the end of each quarter, and will continue to do so.) One group of test developers reduced the number of essays included in their draft instrument by constructing multiple-choice responses using answers students had supplied in their essays. Faculty associated with the two nutrition programs calculated indices of item difficulty and discrimination, and a Cronbach alpha reliability coefficient.

Preparation of Students

All departments decided to require students finishing the core curriculum to take the comprehensive exam. Most often the test was given as part of the senior seminar or capstone course, but performance on the test did not influence the course grade. No minimum score was required for "passing"; students simply were told sometime during the quarter prior to its administration that they would be given an examination designed by the faculty "for the purpose of evaluating and improving curriculum and instruction within the department."

In seven departments students were asked not to study for the exam because the faculty "wanted to see how much they had <u>really</u> retained from their experience in the core curriculum." For one undergraduate exam and for all three at the master's level, students were encouraged to study. The faculty administering the tests reported that most students seemed to be motivated to do their best work; in only two departments did the faculty have any concern that some students may not have taken the test seriously.

Test Administration and Scoring

In six of the eleven departments the comprehensive exam was given during a regularly scheduled senior seminar class or exam period, while for the five others a special time was arranged for all students completing their core course work to take the test. In some cases the exam was administered to graduate students in order to provide some standard against which to assess undergraduate performance, and in one department all faculty took the test. One or two faculty members usually were charged with the responsibility of scoring the test, though several departments used graduate students to assist in this process. Three of the departments used answer sheets to facilitate scoring, three departments that included essay items constructed scoring guidelines to increase the reliability of multiple assessments.

Post-administration Evaluation of the Test

<u>Student reaction</u>. In eight departments students were asked following the experience of taking their comprehensive exam how they felt about it. Most students voiced the opinion that their instructors had developed a difficult test. In one department the students felt that the test was too long and some of the questions were unfair; in another the students resented having to take another senior comprehensive since they had just been required to take the ACT College Outcome Measures Project (COMP) exam in general education in addition to their departmental exam. Students participating in a clinical experience in dietetics were disappointed that their exam did not assess that practical experience. However, the faculty had purposely omitted questions on clinical content because the students in another curriculum within the department had had no clinical component as part of their core curriculum. Advertising seniors were disappointed because their exam did not cover case material taught in one of the core courses.

Four of the departmental exams were pronounced "fair and comprehensive" by students taking them. They felt that the tests had assessed their understanding of most of the important concepts in their major field. The head of the Department of Ornamental Horticulture and Landscape Design routinely interviews seniors before they graduate; in the year following

the first administration of the departmental comprehensive exam, several of the seniors volunteered the information that they thought the comprehensive exam was "a good idea."

Faculty reaction. The entire faculty of the Department of Food Technology and Science took the exam they had developed for students. The department head thought that this experience helped the faculty--many of whom had been trained in just one of a number of specialty areas offered in that department--to understand more fully the nature of the total curriculum. In Advertising the faculty felt that the test was a fair assessment of the students' likelihood of succeeding in the field of advertising. In fact, they felt the test did a better job of arraying students in order by likelihood of success than did the more traditional cumulative grade point average.

Two faculties began with the assumption that the mean score on the departmental exam should be 70 percent of the items correct. In fact, on the majority of the departmental tests the mean percentage correct was between 60 and 65, thus those who were aiming for a 70 percent score were disappointed. Three faculties said "the test was meant to be hard" and were pleased with a student mean score in the range of 65 percent correct. The Dance faculty continued to be frustrated by the fact that without an assessment of students' ability to choreograph and perform a dance, the comprehensive exam covered only half of the core requirements for that curriculum.

Item analysis. In five departments the comprehensive exam was temporarily abandoned soon after it was given because the faculty was immediately inundated by the work connected with redesigning quarter-based courses for the University's proposed conversion to a semester calendar. All planned, however, to give some attention to item analysis before the test was given again. In three departments item difficulty and discrimination indices were calculated following the first administration to students, and one department calculated a Cronbach alpha coefficient of internal consistency. In the Department of Food Technology and Science an analysis was made of the effects on students' scores of their having taken certain courses within the curriculum. The department head in Advertising looked at the relationship between scores on the creative section of the comprehensive exam and cumulative grade point average. He found that there was not a linear relationship: while most students with very high overall GPAs earned high scores on the creative section, some students with rather low GPAs also did well, and for students whose GPAs were not at one of these extremes, there was no discernible relationship between GPA and creative score.

Test revision. Ten of the eleven departmental faculties have not yet made any changes in the comprehensive exam as it was given the first time. In the Food Technology and Science department, several of the essays have been converted to multiple-choice questions since the students' essay responses provided a variety of incorrect answers that could be used as distractors! Most faculties feel that they would benefit by having more students take the same test before they initiate revisions since the number of students taking the test the first time was less than 50.

Test security. Most faculties are not concerned about an immediate need to create new forms of their examinations in order to maintain test security. They perceive that the students who take the test are so close

to graduation that they leave without sharing information about the comprehensive exam with more junior members of the student population. Seven of the departments have a pool of approved and validated items from which to choose, so they can easily create a new form of the test when this becomes necessary. In Advertising, the faculty feels that even if their seniors learn that their exam will contain a creative section, they cannot benefit unduly because this is such a comprehensive exercise that it virtually tests the essence of the advertising curriculum.

Use of Testing Results

The test development process itself had an effect on departments even before the product of that process was administered to students. A common feeling expressed by department heads was that faculty were brought closer together in their thinking about the curriculum as they were forced to focus on common learning objectives for students. Specific changes mentioned by one or more departments included:

- The process has helped the department head enforce consistency among faculty in teaching core courses.
- Faculty are now using the newly developed core competencies in teaching the quarter courses, and will rely upon them again to shape the semester courses for the department.
- A clear progression of courses from lower- to upper-division levels has been established, that is, upper-division courses now actually build upon content students have experienced in lower-division courses.

Two departments took full advantage of their opportunity to work with external consultants in the course of the test development project. The Dance consultant observed classes and provided a brief review of the entire program. The Geography faculty invited their external consultants to give a seminar for faculty and students while on campus to review the comprehensive exam.

Since making the decision to design their own exams, several faculties have discovered that there is an interest in establishing core competencies and/or common exams, perhaps for licensing or registration purposes, within the national organization in their field. Faculty members in four of the eleven departments now are working with their professional associations on competency-writing or test-development projects.

As a result of giving the test to students, most departmental faculties feel they have established a baseline of student performance that will enable them to compare the effectiveness of the semester curriculum with that of the quarter-based curriculum. Moreover, each faculty has identified, as a result of student performance on the test, areas of relative weakness within the curriculum that can be strengthened in the design of courses for the semester system.

Perhaps the most important outcome of faculty studies of student responses on the comprehensive exams is that faculty are now teaching somewhat differently. Most are paying more attention to student experiences that will increase their ability to apply what they're learning in class--providing opportunities for term projects, field trips, and in-class problem-solving. Again as a result of the test development process, instructors are

now more aware of the characteristics of good test items, and have tried to improve the quality of their own course exams.

As a result of student performance on locally developed tests, three faculties decided to change their curriculum requirements; all majors now must take a common core of courses; in the past they were free to select whatever courses they wished from a variety of offerings. In Food Technology and Science a new chemistry series is being required for majors: previously the chemistry series for biology majors had been recommended, but now the faculty believes that the series for chemistry majors would be more appropriate.

Two departments felt that the semester format would improve students' performance on parts of the exam where their scores had been lowest--having the students for a longer period of time under the semester should provide faculty with more opportunities to strengthen students' understanding of a given area.

Finally, two department heads have observed that since their comprehensive exams are based on content of individual courses, the faculty who teach those courses perceive themselves to be evaluated by student scores and are motivated to find ways to improve their teaching effectiveness. "Teaching to the test" is not considered a negative consequence of test development; faculty and department heads say that since their exams were constructed to test essential skills and knowledge, faculty should focus their teaching upon objectives covered on the test.

Future Use of Local Tests

Nine of the eleven departments intend to require their locally developed examinations of every program graduate for the foreseeable future. Four of the programs can make the exam a requirement in a senior seminar or capstone experience. The Geography faculty intends to wait until the semester curriculum is in place before giving its exam again, and the Communications faculty has not yet determined whether its exam will be given again in May 1986 or May 1987.

Four departments have considered giving the senior exam to freshmen for two reasons: (1) to organize the thinking of freshmen about the structure of the major, and (2) to gather baseline data on entering students in order to measure value added over the four-year experience. Two departments are considering giving their test for seniors as a qualifying exam for new graduate students. The undergraduate nutrition examination may be given to juniors in an attempt to diagnose weaknesses that students can work to correct prior to taking the registration exam in dietetics.

Two departmental faculties expressed interest in giving their exam at other universities in order to obtain scores that could be used for comparative purposes--to diagnose relative strengths and weaknesses of the UTK program. No definite plans have been made to do this, however. Finally, one department head remarked that candidates interviewing for faculty positions in his department had expressed interest in looking at the examination for a quick overview of the curriculum as structured by the current UTK faculty.

Faculty Reaction to the Test Development Process

Faculties in five of the eleven departments apparently approached test development with a positive attitude that continued throughout the project. The most frequent complaint from these faculty members was simply that a good deal of time was consumed by the process.

In six departments the initial reaction to having to design a comprehensive exam was somewhat negative: "more paperwork", "busy work from the THEC", "What a waste of time!" were some of the reactions heard by department heads. However, even these negative reactions rather quickly moved to a more positive phase: "If we have to do it, let's do it right," or "Let's use this as an opportunity to do some other things we have wanted to do". As indicated previously, the Dance faculty brought its external consultant to the campus to conduct a mini-program review as well as to review the faculty-developed exam, and the Geography Department used its visiting consultants to provide a seminar for faculty and students.

All faculties now look back at the process and see some benefits. They have a more highly structured core curriculum, and a clearer collective faculty vision of what students should know and be able to do as a result of their work in the major. Many see the need for increasing the students' opportunities to apply what they have learned both in class and in out-of-class experiences, and on course examinations. And finally, they have a baseline of experiences against which to compare the benefits of a semester course format with those of a quarter format.

Advice for Others

Several test developers ended their interview with the writers by offering some advice for other faculties embarking on a test development project. The most often-mentioned warning was to allow plenty of time for the process because it takes longer than most faculty anticipate. Other pieces of advice include:

- Don't use questions from final exams. Most of these are too narrow; coverage should be broader for a comprehensive exam.
- Use a measurement specialist to improve the quality of the items and to conduct item analyses after the test is given.
- Don't underestimate the difficulty of getting all faculty to agree on what should be learned by all students!

One test developer remarked that the faculty time devoted to working on the test is time taken from research--the teaching and advising functions are more immediately demanding and thus receive attention first, while the time for study and research activities is most easily sacrificed. On the other hand, a department head remarked that test development was "an excellent experience--one that every department should use because it focuses on curriculum and instruction in a way that no other exercise can do, and it motivates faculty to correct weaknesses they discover in the process."

Conclusion

The experience of faculty at the University of Tennessee, Knoxville in developing exams in the major field for purposes of assessing and improving

curriculum and instruction has been offered with the hope that it may be instructive for others who are considering a similar endeavor. The test developers themselves will be the first to admit that the technical quality of the instruments they have developed is far from perfect. However, no student is penalized for poor performance on an exam because faculty are not focusing their attention on individual scores. Instead they are looking at mean scores and dispersion of scores about the mean in order to determine strengths and weaknesses of curriculum and instruction. For these purposes the technical quality of individual items is not as significant a factor as it would be if decisions were being made about students on the basis of the results. Finally, the process of test development in many cases has produced benefits for both faculty and students that are independent of student performance on the test itself.

References

Astin, A. W. Achieving Educational Excellence. San Francisco: Jossey-Bass, 1985.

Banta, T. W. "Use of Outcomes Information at the University of Tennessee, Knoxville." In P. T. Ewell (Ed.) Assessing Educational Outcomes. New Directions for Institutional Research, no. 47. San Francisco: Jossey-Bass, September, 1985.

Banta, T. W. and Fisher, H. S. "Performance Funding: Tennessee's Experiment." In J. Folger (Ed.) Financial Incentives for Academic Quality. New Directions for Higher Education, no. 48. San Francisco: Jossey-Bass, December, 1984.

Dressel, P. L. and DeLisle, F. Undergraduate Curriculum Trends. Washington, D.C.: American Council on Education, 1969.

Dressel, P. L., and Associates. Evaluation in Higher Education. Boston: Houghton Mifflin Co., 1961.

Dressel, P. L. Handbook of Academic Evaluation. San Francisco: Jossey-Bass, 1976.

Gordon, L. J. "Comprehensive Examinations in Economics at Denison." AAUP Bulletin, 1958, 54 (September), 622-628.

Jones, E. S. Comprehensive Examinations in American Colleges. New York: MacMillan, 1933.

Levine, A. Handbook on Undergraduate Curriculum. San Francisco: Jossey-Bass, 1978.

Lowell, A. L. "General Exams and Tutors at Harvard College." Educational Record, 1927, 8 (April), 64-67.

Lowell, A. L. "Examination of Subjects Instead of by Courses." Harvard Graduate Magazine, 1912, 20 (30), 584-585.

O'Neill, J. P. "Examinations and Quality Control." In J. R. Warren (Ed.) Meeting the New Demand for Standards. New Directions for Higher Education, no. 43. San Francisco: Jossey-Bass, September, 1983.

Rudolph, F. Curriculum. San Francisco: Jossey-Bass, 1978.

Singletary, O. A. (Ed.) American Universities and Colleges (10th ed) Washington, D.C.: American Council on Education, 1968.

Smallwood, M. L. An Historical Study of Examinations and Grading Systems in Early American Universities. Cambridge: Harvard University Press, 1935.